Available online at http://www.ijasrd.org/in

International Journal of Advanced Scientific Research & Development

Vol. 02, Iss. 03, Ver. II, Jul – Sep' 2015, pp. 135 – 138

e-ISSN: 2395-6089
p-ISSN: 2394-8906

A SURVEY OF OPINION MINING RESEARCH IN HINDI LANGUAGE

Dr. Vivek Kumar Singh

Asst. Prof., Dept., of Computer Science, Banaras Hindu University, Varanasi – 221 005, India.

ARTICLE INFO

Article History:

Received: 20 Sep 2015;
Received in revised form:
23 Sep 2015;
Accepted: 23 Sep 2015;
Published online: 30 Sep 2015.

Key words:

Hindi,
Opinion Mining,
Sentiment Analysis,
Language.

ABSTRACT

This paper presents a literature survey of research work carried on opinion mining in documents in Hindi language. Research output published on sentiment analysis of Hindi language documents is collected. Scopus index is referred to collect publication data. Both journal and conference papers on the topic and indexed in Scopus is collected. The research papers are analyzed to identify important works, levels and methods of sentiment analysis applied to Hindi language documents. The research paper data is for 1998 to 2015. The year-wise publication trend and summary of main research works is presented in paper.

INTRODUCTION

Opinion mining or Sentiment analysis is a language processing analytical method, which enables a user to know the sentiment extremities associated with a sentence or document. On internet data is growing in such a rapid speed that information extraction and analysis has become nightmare. Such huge amount of data requires innovative methods to automatically process the data so that hidden patterns can be realized. Opinion mining gives chance to a service provider or a brand to measure its strength through word of mouth.

It is observed that more than 50% of the web based systems are developed in English language. Thus most information processing and analysis related advance techniques are developed for English language. It can be observed that Hindi languages presence is exponentially increasing over the web. People have started developing blogs, forums, product support sites, review engines and online literature banks in Hindi language. Hindi language has its own strengths so methods developed for English language either needed to be tuned or re-developed to deal with Hindi documents. It is high time to develop some text analytical methods for Hindi so that a hidden treasure of web can be explored. This paper tries to survey the research work done on sentiment analysis in Hindi language. A scientometric approach is used to collect and analyze data.

How to cite this article: Singh, V. K., (2015). "A Survey of Opinion Mining Research in Hindi Language". *International Journal of Advanced Scientific Research & Development (IJASRD), 02* (03/II), pp. 135 – 138.

SENTIMENT ANALYSIS RESEARCH IN HINDI

In the 2001 Indian census, 258 million (258,000,000) people in India reported Hindi to be their native language. There are many websites which provide information in Hindi such as news sites: http://dir.hinkhoj.com/, http://bbc.co.uk/hindi, sites which have information about culture, music, entertainment and other aspects of arts: http://www.webdunia.com/, http://www.virarjun.com/, http://www.raftaar.in/ etc. Availability of Hindi text in social media attracted researchers to contribute their work towards it. Relatively, a decent amount of work has been done on Hindi text in recent years. A brief summary of sentiment analysis task on Hindi text is as follows:

Sentiment Analysis / Opinion Mining: The work done by Lahiri (1998) is one of the pioneering works on sentiment analysis or opinion mining. This paper presented an analysis of negative polarity items (NPI) in Hindi by applying some rules. This work aimed to show the meaning of NPI's in Hindi and the reasons for their behavior as NPI. Finally, a comparison with analyses of English is also provided. The book written by Kumar (2006) also discusses about handling Negative Polarity Items (NPI) in Hindi.

Mogadala and Varma (2012b) have done the task of extracting opinion about people (opinion target) in Hindi news articles. This approach extracts opinion words from English collection of comparable corpora to get transliterated and translated to Hindi languages. Transformed opinion words are then used to create subjective language model (SLM) and structured opinion queries (OQs) using inference network (IN) for retrieval to confirm the opinion about opinion targets in documents. In another work, Mittal et.al. (2013) have done discourse based i.e. topic based sentiment analysis on Hindi reviews. In this paper, it is also investigated that how by proper handling of negation and discourse relation may improve the performance of Hindi review sentiment analysis

In another paper, Patra et.al. (2013) proposed an unsupervised method to classify music by mood. In this study, a fuzzy c-means classifier is used to do the automatic mood classification. The dataset consists of 250 Hindi songs are used in this experimental study. Later, Choudhary et.al. (2014) proposed a rule-based methodology for identification of tense, aspect and mood (TAM) features in a given Hindi text.

Subjectivity Analysis: Mogadala and Varma (2012a) proposed sentence level subjectivity classification using Entropy based category coverage difference criterion (ECCD) feature selection method and language independent feature weighing method which are consistent across languages. Experiments are performed on five different languages including Hindi, English, Romanian, French and Arabic. MPQA corpus is used for English, Romanian, French and Arabic language experiments. Hindi experiments were performed using sentences from the news corpus.

Sentiment Lexicon Development: Rao and Ravichandran (2009) focused a semi-supervised learning framework for building sentiment lexicons in a variety of resource availability situations. In this work, polarity detection is treated as a semi-supervised label propagation problem in a graph. The evaluation is done with Hindi, English and French datasets.

Improving Word Senses: The work proposed by Jain and Lobiyal (2014) updated the word senses in Hindi WordNet. This paper proposed a graph based model and its associated techniques to automatically acquire words' senses.

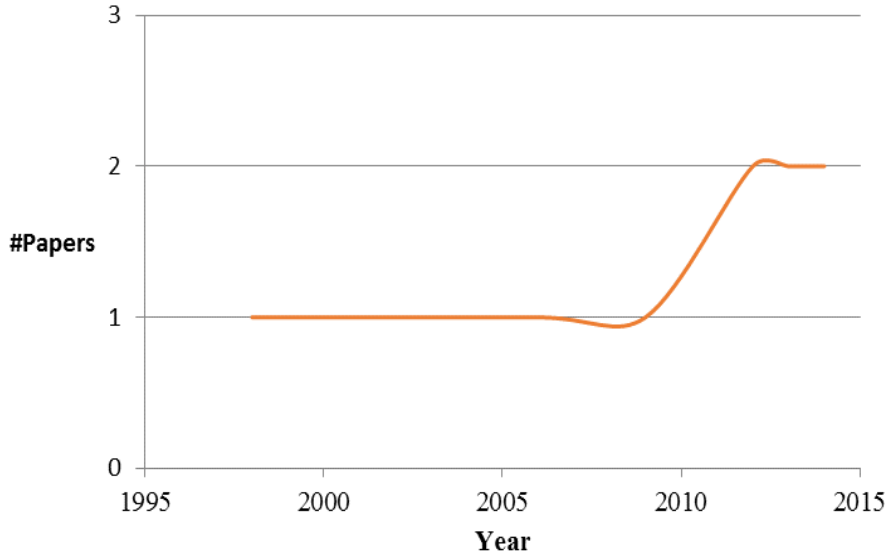
Table 1: *Dataset in Hindi Language*

S. No.	Name	No. of Papers
1	News	2
2	Reviews	1
3	Music Clips	1

Table 2: *Sentiment Analysis Research in Hindi*

Topic	Method	Author	Dataset
Sentiment Analysis/ Handling Negative Polarity Items/ Discourse Based Sentiment Analysis/ Mood Extraction	Rules	Lahiri (1998)	
		Kumar (2006)	
	opinion queries (OQs) using inference network (IN)	Mogadala and Varma (2012b)	News
		Mittal et.al. (2013)	Reviews
	fuzzy c-means classifier	Patra et.al. (2013)	Music clips
	rule-based method	Choudhary et.al. (2014)	
sentence level subjectivity classification	Entropy based category coverage difference criterion (ECCD), feature selection method and language independent feature weighing method	Mogadala and Varma (2012a)	Hindi-News; English, Romanian, French and Arabic - MPQA
Polarity lexicon induction	semi-supervised label propagation	Rao and Ravichandran (2009)	
Improving word senses in HindiWordNet	graph based model	Jain and Lobiyal (2014)	

Figure 1: *Year-wise Publication Plot*



CONCLUSION

This paper presents a survey of sentiment analysis research work done in Hindi language. Research output published in journals and conferences indexed in Scopus on the topic of opinion mining/ sentiment analysis is collected and analyzed. It is seen that research work on opinion mining and sentiment analysis in Hindi has already begun and different researchers carried out work on development of tools and algorithms for sentiment analysis in Hindi. The paper presents an informative account of sentiment analysis research work in Hindi.

REFERENCES

[1] Choudhary, N., Pandey, P., & Jha, G. N. (2014). A Rule based Method for the Identification of TAM features in a POS Tagged Corpus. In Human Language Technology Challenges for Computer Science and Linguistics, pp. 178-188, Springer International Publishing.

[2] Jain, A., & Lobiyal, D. K. (2014, February). A new method for updating word senses in hindi wordnet. In Proccedings of International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), pp. 666-671.

[3] Kumar, R. (2006). Negation and licensing of negative polarity items in Hindi syntax. Taylor & Francis.

[4] Lahiri, U. (1998). Focus and negative polarity in Hindi. Natural Language Semantics, 6(1), 57-123.

[5] Mittal, N., Agarwal, B., Chouhan, G., Pareek, P., &Bania, N. (2013). Discourse Based Sentiment Analysis for Hindi Reviews. In Pattern Recognition and Machine Intelligence, pp. 720-725, Springer Berlin Heidelberg.

[6] Mogadala, A., & Varma, V. (2012a). Language Independent Sentence-Level Subjectivity Analysis with Feature Selection. In 26th Pacific Asia Conference on Language, Information and Computation (PACLIC 26).

[7] Mogadala, A., & Varma, V. (2012b). Retrieval approach to extract opinions about people from resource scarce language News articles. In Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, pp. 4, ACM.

[8] Patra, B. G., Das, D., & Bandyopadhyay, S. (2013). Unsupervised approach to Hindi music mood classification. In Mining Intelligence and Knowledge Exploration, pp. 62-69, Springer International Publishing.

[9] Rao, D., & Ravichandran, D. (2009). Semi-supervised polarity lexicon induction. In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (pp. 675-682). Association for Computational Linguistics.